

Weekly Report

05/18/2015 - 05/24/2015

Jing XIA

May 25, 2015

1 Summary

This week I mainly focus on the DataScanner2 project (the second phase of data inspection).

2 Projects

2.1 Project 1 - DataScanner2

DataScanner2: A system that integrates various data quality metrics, covering all quality metrics categories and providing a quality-aware (clustering, correlation, outliers, complex patterns, image quality, feature preservation) visualization.

Because of bad results from Scagnostics, I've implemented SIFT similarity of view images. However, SIFT is not good either (see Figure 1 below). The similarities to figure (a) are 0.02, 0.42, 0.64 respectively, but the degree of similarity does not correspond to their similarity values. It seems SIFT consider two figure more similar when the spacing between dots are similar. Makes sense for images. So maybe we should turn to dot distribution instead of SIFT...

The first thing to do next is to modify the original quartet method (altering the parameters or local process maybe) for a better result. The second thing is to quantify each view with quality metrics and visualize it in the web interface. I'm also trying to contribute other graph-based features to views according to identify features from the following views (figures attached separately because of file size limits).

2.2 Project 2 - NBA Game Visualization

-

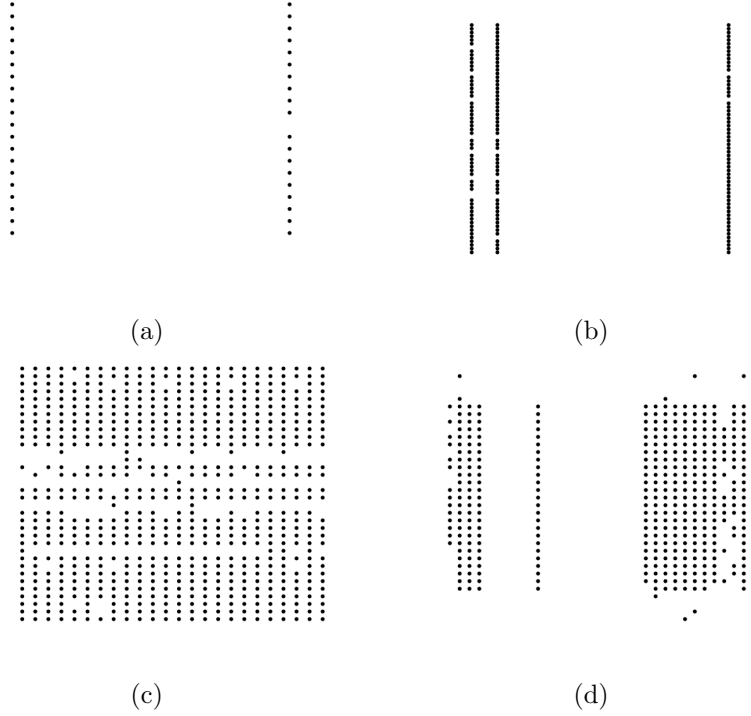


Figure 1: The similarities to figure (a) are 0.02, 0.42, 0.64 respectively.

2.3 Project 3 - FlowTracker

We sent emails to the dataset provider for answers and got one reply but have not figured out the problem yet. We also sought for game dataset for the finance condition of the players and thus the system.

3 Paper Reading

-

4 To Do List

1. DataScanner2 — view-wise quality metric implementation.
2. DataScanner2 — web interface implementation.
3. FlowTracker — New dataset.